A Class of Mixtures with Application to Clinical Chemistry and Nutritional Survey Data

CAVELL BROWNIE, JEAN-PIERRE HABICHT and DOUGLAS S. ROBSON*

Mixtures of normal distributions have been used in modeling a wide variety of experimental data in fields as different as fisheries biology, population genetics, medicine and physics. A problem in the field of nutrition concerning estimation of the population proportion of anemic individuals from a sample of hemoglobin values has led to consideration of a class of mixtures of two components, where only the predominant component is assumed to be normal. The second component is assumed to contaminate one side only of the normal component (the side of contamination being known) but is of otherwise unspecified form. An estimation procedure is outlined for this problem. Monte Carlo methods are used to compare this procedure with the maximum likelihood (ML) method for mixtures of two normal components when the true model is (a) a Beta-normal mixture and (b) a normal-normal mixture. This involves examining the performance of the much-discussed ML method in situations not previously studied. Results from the application of the two methods to real data are described.

KEY WORDS: Anemia standards; Mixed distributions; Normal mixtures; Maximum likelihood; Percentile estimates.

*Cavell Brownie is Assistant Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Her research was supported by USDA Competitive Research Grant No. 5901-0410-9-0314-0 while at the Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853. Jean-Pierre Habicht is Professor, Division of Nutritional Sciences and Douglas S. Robson is Professor, Biometrics Unit, Cornell University, Ithaca, NY 14853.

1. INTRODUCTION

The cumulative distribution function (cdf) $F(x)$ is called a finite distribution mixture if it is of the form

$$F(x) = \sum_{i=1}^{k} p_i F_i(x) \quad ,$$

where

$$0 < p_i < 1, \qquad i = 1, \cdots, k, \qquad \sum_{i=1}^{k} p_i = 1 \quad ,$$

$$k > 1 \text{ is a finite integer} \quad ,$$

$$\text{and the } F_i \text{ are each proper cdf's} \quad . \tag{1}$$

The p_i are called the mixing proportions and the F_i are called the component distributions. When the components F_i in (1) are each normal [i.e., F_i is the $N(\mu_i, \sigma_i^2)$ cdf given by $F_i(x) = \int_{-\infty}^{x} (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x - \mu_i)^2/\sigma_i^2\}dx$], the resulting mixture is called a normal mixture or a mixed normal distribution. Normal mixtures have been used to model a wide variety of practical problems. Examples of such applications are given by Hosmer (1973a) and MacDonald and Pitcher (1979), fisheries biology; Elston et al. (1974), population genetics; Clark et al. (1968), medicine; and Gindler (1970), clinical chemistry.

There is a large literature, dating back to Pearson (1894) concerning the problem of estimating the parameters of a normal mixture; for example, see the article by Quandt and Ramsey (1978) and the comments following by Hartley (1978), Hosmer (1978), Kiefer (1978a), Binder (1978), Fowlkes (1978), Bryant (1978), Clarke and Heathcote (1978), and Johnson (1978). We briefly review some of the results relating to the method of maximum likelihood (ML) in Section 5.

The problem of estimating the prevalence of anemia from a sample of hemoglobin (Hb) values has led to consideration of a class of mixtures of two

components, the predominant component being normal and contaminated on one side only by a component of unspecified parametric form, the side of contamination being known. Subject matter considerations leading to this model and a formulation of the statistical problem appear in Section 2. A procedure for estimating the mixing proportions and parameters of the predominant normal component is described in Section 3.

The performance of this procedure is compared with that of the ML method for mixtures of two normal components, with emphasis on the situation where the proportion of contamination is small. Comparisons are made using Monte Carlo methods for situation where the contaminating group or secondary component is (a) Beta-distributed (Section 5) and (b) also normally distributed (Section 6). Thus results for the performance of the ML method are presented for situations which do not appear to have been studied previously. Section 7 contains a discussion of the Monte Carlo results, and Section 8 contains results from applying both methods to data from a nutritional survey.

2. MOTIVATION AND DEFINITION OF THE PROBLEM

Anemia is traditionally defined in terms of blood Hb concentrations. For example, the WHO (1968) criterion for nonpregnant women defines anemia as having an Hb level of less than 12 g/dl. The traditional method of estimating the proportion or prevalence of anemic women in a population is by the sample fraction with Hb values less than 12. As Hb distributions for anemics and non-anemics overlap (Cook et al., 1971, Garby et al., 1969), estimating the prevalence of anemia in this way ignores errors of classification, and will usually be biased. Recognizing this, Cook et al. (1971) used a mixture of two normal components to model the distribution of Hb values in a population containing anemics and non-anemics in unknown proportions. Meyers (1978) questioned the validity of the normality assumption for the anemic Hb distribution and proposed a model

with a predominant non-anemic normally distributed component, contaminated in the lower tail only by an anemic component of unspecified parametric form and present in an unknown proportion.

Becktel (1970) describes a similar problem in the related field of clinical chemistry, the only difference being that estimation of the proportion of the contaminating component was not of interest. Martin et al. (1975) devote a chapter of their text for clinical chemists to the use of mixtures in analyzing laboratory test results. They also argue that normality may not hold for the secondary component. They propose approximating the functional form of the component densities by Gram-Charlier series. Such curve fitting procedures based on the substitution of arbitrary functional forms are unappealing as they are not motivated by, or descriptive of, the underlying biological problem, and are not discussed further here. We take an alternative approach to the problem as described below.

The distribution of Hb values in the situation described by Meyers (1978) can be represented by the mixture cdf

$$F(x) = pA(x) + (1-p)G(x) \qquad (2)$$

where

(i) $0 < p < 1$,

(ii) $G(x)$ is the $N(\mu, \sigma^2)$ cdf ,

(iii) $A(x)$ is a continuous cdf with density $a(x)$, and

(iv) There is some unknown x_0 such that $x_0 < \mu$ and $A(x) = 1$ for $x \geq x_0$.

Note that the assumption (iv) above is one possible representation of the statement that contamination occurs on only one side of the normal component

(G). As the side of contamination is assumed known, it is taken here without loss of generality to be the lower side (i.e., $A(x) = 1$ for all $x \geq \mu$). Without (iv), the class of mixtures generated by (2) is not identifiable. Although (2) is a biologically reasonable model, because of (iv) it does not include as a special case the commonly used mixture of two normal components. For convenience, the mixture of two normals is referred to here as a normal-normal mixture.

Given the model (2), we need a procedure for obtaining estimates of μ and σ^2 (the mean and variance of the normal component G), and p (the proportion of the non-normal component A).

3. ESTIMATION OF μ, σ AND p

Meyers (1978) and Becktel (1970) proposed estimation procedures which are dependent on the assumption that p is "small". We outline below an iterative estimation procedure which is also appropriate for p small. This procedure is compared briefly with the methods of Meyers and Becktel in Section 4.

3.1 Basis of the Method

The following results are made use of in the estimation of μ, σ and p. From (2), part (iv):

$$F(x) = p + (1-p)G(x) \quad \text{for } x \geq x_0 \ ,$$

hence

$$F(\mu) = p + (1-p)(\cdot 5) \doteq 0.5 \quad \text{for p small} \tag{3.1}$$

$$F(\mu+\sigma) = p + (1-p)(\cdot 8413) \doteq 0.84 \quad \text{for p small} \ . \tag{3.2}$$

Estimation of p is based on the function $\lambda(x)$ defined by

$$\lambda(x) = \frac{F(x)-G(x)}{1-G(x)} = p\frac{A(x)-G(x)}{1-G(x)} \leq p = \frac{F(x)-G(x)}{A(x)-G(x)} \ , \tag{3.3}$$

where

$$\lambda(x)\begin{cases} < p \text{ for } x < x_0 \\ = p \text{ for } x \geq x_0 \end{cases} \ . \tag{3.4}$$

A naive estimator of $\lambda(x)$, given estimates $\hat{\mu}$, $\hat{\sigma}$ of μ, σ, is

$$\hat{\lambda}(x) = \frac{\hat{F}(x)-\hat{G}(x)}{1-\hat{G}(x)} = \frac{\hat{F}(x)-\Phi\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)}{1-\Phi\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)} \quad , \tag{3.5}$$

where

$\hat{F}(x)$ is the value of the empirical or sample cdf at the point x ,

and

Φ is the standard normal cdf .

Estimation of p using (3.4) and (3.5) is not straightforward because

(i) x_0 is unknown,

(ii) $\hat{\lambda}(x)$ is not unbiased as an estimator of $\lambda(x)$, and the bias depends on x,

(iii) the $\hat{\lambda}(x)$ are highly correlated for adjacent x-values, and

(iv) $\mathrm{Var}[\hat{\lambda}(x)]$ is roughly proportional to $[1 - G(x)]^{-2}$ and increases rapidly as x and $G(x)$ increase.

How to determine an optimal rule for estimation of p from $\hat{\lambda}(x)$ remains an open question. However, a heuristic account of some aspects of the problem and related empirical results are given here.

From (3.4) we know that as x approaches x_0 from below, $\lambda(x)$ will approach p from below. We assume $\hat{\lambda}(x)$ is based on well behaved estimators $\hat{\mu}$ and $\hat{\sigma}$. For given A and G, we can find a set of x-values $\{x_i ;\ i = 1, \cdots, \ell,\ x_1 < x_2 < \cdots < x_\ell\}$, with corresponding variables $\{\hat{\lambda}(x_i),\ i = 1, \cdots, \ell\}$, such that the event $\hat{\lambda}(x_{i-1}) > \hat{\lambda}(x_i)$ has greater probability if $x_i > x_0$ than if $x_i < x_0$. That is, $P[\hat{\lambda}(x_{i-1}) > \hat{\lambda}(x_i)]$ will increase sharply as x_i increases beyond x_0.

Given the set $\{x_i,\ i = 1, \cdots, \ell\}$ a rule, called the first reversal rule, which will determine $\hat{p}$ from $\{\hat{\lambda}(x_i),\ i = 1, \cdots, \ell\}$, sets

$$\hat{p} = \hat{p}_{RR1}[\{\hat{\lambda}(x_i)\}] = \tfrac{1}{2}[\hat{\lambda}(x_{m-1}) + \hat{\lambda}(x_m)]$$

where (3.6)

$$x_m = \min_{2 \le i \le \ell} \{x_i ;\ \hat{\lambda}(x_{i-1}) > \hat{\lambda}(x_i)\} \quad .$$

The first reversal rule (RR1) is just one of many rules which can be used to obtain $\hat{p}$ from $\{\hat{\lambda}(x_i)\}$. The properties of such a rule will of course depend on the set $\{x_i\}$. There are as yet no results concerning sets $\{x_i\}$ and rules which have optimal properties for general F as defined in (2). Practical suggestions concerning the choice of $\{x_i\}$ are made in Section 4, and empirical results are presented in Sections 5, 6, and 8.

A serious objection to the above method is that for a given set $\{x_i\}$, monotonicity of the corresponding $\{\lambda(x_i)\}$ may not hold for all A and G. Monotonicity holds under the following conditions: Let a(x), g(x) represent the densities corresponding to A(x) and G(x). Then for $x < x_0$, $\lambda(x)$ is increasing (respectively decreasing) as the Mills' ratio (Kendall and Stuart, 1977, p. 144) $[1 - G(x)]/g(x)$ is $>$ (respectively $<$) $[1 - A(x)]/a(x)$. The decreasing Mills' ratio for the normal component G must exceed 1.253σ for $x < x_0 < \mu$ while the Mills' ratio for A may be expected to approach zero at x_0, and calculations for various A and G suggest that the range of x-values for which $\lambda(x)$ is decreasing do not pose a serious problem. In the anemia problem the question is whether the ratio $P[Hb > x|\text{non-anemic}]/P[Hb > x|\text{anemic}]$ is increasing for $x \le x_0$, as must certainly be the case in some interval $a < x \le x_0$. (Situations where monotonicity of $\lambda(x)$ may break down do, however, need to be investigated further.)

3.2 The Iterative Percentile Method (IPM) for Estimation of μ, σ, p

We now outline an iterative estimation scheme, based on the results in Section 3.1:

Let X_i, $i = 1, \ldots, n$ be iid r.v.'s with cdf F as in (2), and let

$\hat{X}_r$ be the 100r percentile $(0 < r < 1)$ of the sample $X_1, \ldots, X_n$.

The Iterative Percentile Method (IPM)

Step 0

(a) Obtain initial estimates of μ, σ, G, λ as follows:

$$\hat{\mu}^0 = \hat{X}_{.5000} \qquad \text{from (3.1)} ,$$

$$\hat{\sigma}^0 = \hat{X}_{.8413} - \hat{\mu}^0 \qquad \text{from (3.2)} ,$$

$$\hat{G}^0(x) = \Phi\left(\frac{x-\hat{\mu}^0}{\hat{\sigma}^0}\right) .$$

For the predetermined x-values $\{x_i;\ x_1 < x_2 < \cdots < x_\ell\}$,

$$\hat{\lambda}^0(x_i) = \frac{\hat{F}(x_i)-\hat{G}^0(x_i)}{1-\hat{G}^0(x_i)} , \qquad i = 1, \cdots, \ell .$$

(b) Apply a rule (e.g., RR1) to $\{\hat{\lambda}^0(x_i)\}$ to obtain an initial estimate of p.

Thus

$$\hat{p}^0 = \hat{p}_{RR1}[\{\hat{\lambda}^0(x_i)\}] = \tfrac{1}{2}[\hat{\lambda}^0(x^0_{m-1}) + \hat{\lambda}^0(x^0_m)]$$

where

$$x^0_m = \min_{2 \le i \le \ell} \{x_i;\ \hat{\lambda}^0(x_{i-1}) > \hat{\lambda}^0(x_i)\} .$$

Step ν, $\nu = 1, 2, \cdots$

At step ν, use the estimate $\hat{p}^{\nu-1}$ from step ν-1 to obtain $\hat{\mu}^\nu$, $\hat{\sigma}^\nu$, $\hat{p}^\nu$ by

$$\hat{\mu}^\nu = \hat{X}_r \qquad \text{where } r = \hat{p}^{\nu-1} + (1 - \hat{p}^{\nu-1})(\cdot 5) ,$$

$$\hat{\sigma}^\nu = \hat{X}_s - \hat{\mu}^\nu \qquad \text{where } s = \hat{p}^{\nu-1} + (1 - \hat{p}^{\nu-1})(\cdot 8413) ,$$

$$\hat{p}^\nu = \hat{p}_{RR1}[\{\hat{\lambda}^\nu(x_i)\}] = \tfrac{1}{2}[\hat{\lambda}^\nu(x^\nu_{m-1}) + \hat{\lambda}^\nu(x^\nu_m)] ,$$

with x^ν_m defined with respect to $\{\hat{\lambda}^\nu(x_i)\}$ as in (3.6).

Determine whether convergence has occurred, i.e., if $|\hat{\mu}^\nu - \hat{\mu}^{\nu-1}| \le K_1$, $|\hat{\sigma}^\nu - \hat{\sigma}^{\nu-1}| \le K_2$, and $|\hat{p}^\nu - \hat{p}^{\nu-1}| \le K_3$, where K_i, $i = 1, 2, 3$, are arbitrary predetermined constants. If convergence has occurred, then $\hat{p} = \hat{p}^{\nu-1}$, $\hat{\mu} = \hat{\mu}^{\nu-1}$, $\hat{\sigma} = \hat{\sigma}^{\nu-1}$. Otherwise, proceed to Step $\nu+1$.

3.3 Nonconvergence of the IPM

For a given data set, the IPM may not converge. Nonconvergence is declared in the following situations:

(i) When for some $\nu \geq 0$, the RR1 does not produce an estimate $\hat{p}^{\nu}$, because $\hat{\lambda}^{\nu}(x_{i-1}) < \hat{\lambda}^{\nu}(x_i)$, $i = 2, \cdots, \ell$.

(ii) When for some $\nu \geq 0$, x_m^{ν} is such that $\hat{G}^{\nu}(x_m^{\nu})$ is arbitrarily large (say $> .84$). This is because $\mathrm{Var}[\hat{\lambda}(x)]$ increases rapidly as $G(x)$ increases (Section 3.1), so that $\hat{\lambda}(x)$ becomes increasingly unreliable.

(iii) When $\hat{p}^{\nu} < 0$ for some $\nu \geq 0$.

4. PROPERTIES OF THE IPM

4.1 Comparison with the Methods of Meyers and Becktel

The estimates of μ and σ obtained graphically by Meyers (1978) are analogous to the initial estimates $\hat{\mu}^0$ and $\hat{\sigma}^0$ of Step 0 of the IPM, but her method for obtaining $\hat{p}$ is not clearly defined. Meyers' estimates of μ and σ have positive and negative bias, respectively, resulting in an underestimate of p. This bias will be non-neglible if p is greater than 3-5 percent.

The method Becktel (1970) is also dependent on p small, and bears some similarity to Step 0 of the IPM. It differs in that $\hat{p}$ is obtained from $\hat{\mu}$ in a manner analogous to setting $\hat{p} = \hat{\lambda}(\hat{\mu}) = [\hat{F}(\hat{\mu}) - .5]/(1 - .5) = 2[\hat{F}(\hat{\mu}) - .5]$. Then $\hat{\sigma}$ is obtained from a linear combination of $\hat{\mu}$ and six adjusted (to allow for $\hat{p}$) sample percentiles.

When p is small, Step 0 of the IPM and Meyers' method will give similar results. Becktel's method, though similar in spirit, depends on $\hat{\lambda}(\hat{\mu})$, and as $\mathrm{Var}[\hat{\lambda}(\hat{\mu})]$ will usually be larger than $\mathrm{Var}[\hat{\lambda}(x_m^0)]$ of the IPM, Bectel's method is likely to produce more variable results. When p is not small the IPM will be less biased than Meyers' and Becktel's methods.

4.2 Inefficiency of the Percentile Estimates of μ and σ

An obvious criticism of the IPM is that it is based on inefficient percentile estimators. Hence large samples ($n > 300$ in most situations) are needed for reasonable accuracy. More efficient estimates of μ and σ might be obtained at each iteration in the following way. At the νth iteration, $\nu \geq 1$, assume $x_m^{\nu-1}$ determines a truncation point to the right of which there is no contamination, i.e., assume

$$P[X \leq x | X > x_m^{\nu-1}] = \frac{G(x)-G(x_m^{\nu-1})}{1-G(x_m^{\nu-1})}, \qquad x_m^{\nu-1} \leq x \leq \infty \quad , \tag{4.1}$$

where (4.1) is a truncated normal distribution. Then $\hat{\mu}^{\nu}$, $\hat{\sigma}^{\nu}$ would be obtained from observations to the right of $x_m^{\nu-1}$ by ML, using (4.1) as described by Cohen (1950). The properties of this alternative to the computationally simpler IPM have not been investigated empirically.

4.3 Properties of the IPM Estimates

Analytical results concerning properties of the IPM estimators are not easily obtained due to the distribution-free nature of the component $A(x)$. However, large sample variances have been derived for $\hat{G}(x)$ and $\hat{\lambda}(x)$. The appropriate formulae are contained in the appendix. These have been shown empirically to be valid when p is small and the procedure does not usually iterate, i.e., go beyond Step 0. For larger p, when iteration will usually occur, these asymptotic variances will not be appropriate for sample sizes usually met with in practice ($n < 4{,}000$, say).

$\mathrm{Var}(\hat{p}_{RRl})$ depends on the distribution of $\hat{\lambda}(Y_m)$ where Y_m is a r.v. with outcome corresponding to the value of x_m^{ν} at convergence. Thus

$$\mathrm{Var}(\hat{p}_{RRl}) = \mathrm{Var}_{Y_m}\left\{E[\tfrac{1}{2}\hat{\lambda}(x_{m-1}) + \tfrac{1}{2}\hat{\lambda}(x_m) | Y_m = x_m]\right\}$$

$$+ E_{Y_m}\left\{\mathrm{Var}[\tfrac{1}{2}\hat{\lambda}(x_{m-1}) + \tfrac{1}{2}\hat{\lambda}(x_m) | Y_m = x_m]\right\} \quad . \tag{4.2}$$

If the distribution of Y_m is concentrated on a subset of $\{x_i\}$ containing a few elements only, then (4.2) can be used in estimating $\mathrm{Var}(\hat{p}_{RR1})$ as the first term will be negligible. This is not the case in the situations studied in Sections 5 and 6.

For RR1 and similar rules, properties of the estimators will in general improve as the separation between the two components increases, since as $|\mu - x_0|$ increases, $G(x_0)$ and $\mathrm{Var}[\hat{\lambda}(x_0)]$ decrease.

Empirical results for the performance of the IPM are presented in Sections 5 and 6.

4.4 Convergence Properties

At Step 0, p is treated as negligible, so that $\hat{\mu}^0$ underestimates μ, and $\hat{\sigma}^0$ overestimates σ. As $\hat{p}^\nu > 0$, $\nu \geq 0$ [see Section 3.3 (iii)], the estimates $\hat{\mu}^\nu$ are adjusted upwards, $\hat{\sigma}^\nu$ adjusted downwards, during iteration. Convergence will usually occur for $\hat{\mu}^\nu$ and $\hat{\sigma}^\nu$ close to μ and σ, respectively. If iteration results in an overshoot (i.e., $\hat{\mu}^\nu > \mu$, $\hat{\sigma}^\nu < \sigma$), nonconvergence may occur as successive iterations produce increasingly larger $\hat{p}^\nu$, $\hat{\mu}^\nu$ and increasingly smaller $\hat{\sigma}^\nu$. In this situation x_m^ν moves to the right and $\hat{G}(x_m^\nu)$ increases correspondingly, resulting in nonconvergence as described in Section 3.3 (ii).

A second reason for nonconvergence is $\hat{p}^\nu < 0$, for some $\nu \geq 0$, [Section 3.3 (iii)]. This may occur at Step 0 for p very small, or for p moderately large (> 20%) when $\hat{\mu}^0$ and $\hat{\sigma}^0$ are such that $\hat{G}^0(x)$ is grossly overestimated for $x < x_0$.

The probability of convergence will depend on the criteria K_1, K_2, K_3. For slightly discrete data, such as Hb values which are commonly reported to accuracy of about $(1/10)\sigma$, $K_1 = K_2 = K_3 = 0$ can be used. Empirical results in Sections 5 and 6 show satisfactory convergence ratios for several rules using $K_1 = K_2 = K_3 = 0$.

Convergence will also be influenced by the size of the increments $(x_i - x_{i-1})$. In general, the larger the increments, the smaller the probability of convergence.

4.5 Choosing the $\{x_i\}$

In the absence of information concerning x_0 or the form of A, it seems reasonable to choose equally spaced x_i, $i = 1, \cdots, \ell$. The values of x_i, $(x_i - x_{i-1})$ and ℓ must then be determined. The value of ℓ can be chosen so that x_ℓ lies at the upper end of the range of the observations because of the check contained in Section 3.3 (ii). An alternative to using subjective information in specifying x_1 is to choose x_1 near the lower end of the range of sample values and adopt the convention that $\hat{\lambda}(x_i) < \hat{\lambda}(x_{i-1})$ is regarded as a reversal only if $\hat{F}(x_i) > \hat{F}(x_{i-1})$. This convention, with x_1 as the 0.5 percentile of the sample, was used in the computer routine in the empirical studies in Sections 5 and 6. Information concerning the approximate value of p (and np) will influence the value of $(x_i - x_{i-1})$. In general for np small, $(x_i - x_{i-1})$ should be small also. Frequently, in practice, an approximate value for σ is known, and empirical results in Sections 5 and 6 show that an increment of $(x_i - x_{i-1}) = \sigma/10$ will often work. For large samples ($n > 500$) and $p \leq 3\%$, smaller increments may be used, and for larger p, say $p \geq 10\%$, $\sigma/5$ may be appropriate.

4.6 Why Use the IPM?

In developing the IPM, motivation was provided by predominance of the normal component G (corresponding to small p), and the nonnormality of A. Even if A is markedly nonnormal, when p is small the mixture of two normals may be a reasonable approximation and a procedure developed for normal-normal mixtures may give better (e.g., w.r.t. mean squared error) results than the IPM. Martin et al. (1975), pp. 325-327, present an artificial example which shows that the graphical method of Neumann (1968) for normal-normal mixtures may be biased in the case of skewed A and normal G. However, the performance of an analytical

method such as ML does not appear to have been studied in such a situation. We have therefore investigated empirically the question of whether situations exist for which the IPM is an appropriate procedure.

5. COMPARISON OF THE IPM AND NNML FOR A BETA-NORMAL MIXTURE

5.1 The NNML Method

The ML method is one of the most widely used and probably the most controversial of the methods for normal-normal mixtures. We will refer to this here as the NNML method to emphasize that the likelihood on which the ML procedure is based is that of the normal-normal mixture.

Kiefer and Wolfowitz (1956) showed that the NNML estimators were not consistent if the variances of the two components differed ($\sigma_1^2 \neq \sigma_2^2$). Other authors have stated that NNML "breaks down" or is inconsistent if $\sigma_1 \neq \sigma_2$, e.g., Day (1969), Fryer and Robertson (1972), Quandt and Ramsey (1978). Hasselblad attempted to side-step this issue by proposing ML for grouped data, the likelihood in this situation being multinomial.

In contrast to the theoretical results of Kiefer and Wolfowitz (1956), various simulation studies showed that NNML for grouped and ungrouped data has reasonable properties when $\sigma_1 \neq \sigma_2$ if there is reasonable separation between the component distributions, e.g., Hasselblad (1966), Dick and Bowden (1973), Quandt (1972), Fowlkes (1979). The NNML method is said to be superior to the method of moments both when $\sigma_1 = \sigma_2$, e.g., Day (1969), Tan and Chang (1972), and when $\sigma_1 \neq \sigma_2$, e.g., Fryer and Robertson (1972). N. M. Kiefer (1978b) attempted to resolve the discrepancy between theoretical and empirical results showing that if the parameter space is suitably defined, for large enough n, the likelihood equations have a consistent root.

Fowlkes (1979) showed empirically for grouped and ungrouped data and $\sigma_1 \neq \sigma_2$, that NNML was superior in terms of mean squared error to a least squares

procedure. Methods using empirical moment generating functions, e.g. Quandt and Ramsey (1978), and empirical characteristic functions, e.g., Clarke and Heathcote (1978), have not yet been widely studied.

In view of the above, we chose to compare performance of the IPM with that of the NNML method. These two methods are not strictly comparable, as NNML estimates a greater number of parameters but uses more information than the IPM. Nevertheless, we can compare certain aspects of performance which are important to the estimation problem described in Section 2.

5.2 A Monte Carlo Study with F a Beta-Normal Mixture

The comparison of the IPM and NNML method when (2) holds involves examining the robustness of NNML to departures from normality in A. Situations where we might expect NNML to be non-robust arise when A is markedly skewed with a long lower and short upper tail.

The distributions A and G of (2) used in the Monte Carlo study are as follows:

G is the N(0,1) cdf, giving $\mu = 0$ and $\sigma = 1$

Let Z have the Beta (α, β) distribution on (0,1) with $\alpha = 1.2735$ and $\beta = 1.839443$

Then $X = -3.3Z - 1.35$ has distribution A with mean $\mu_A = -2.7$, mode -2.16, variance 0.64, standard deviation 0.80 and density

$$a(x) = \begin{cases} (3.3)^{1-\alpha-\beta}(-1.35-x)^{\alpha-1}(4.65+x)^{\beta-1}/B(\alpha,\beta) & -4.65 \le x \le -1.35 \\ 0 \quad \text{otherwise} \end{cases} .$$

Thus $x_0 = -1.35$, and the resulting F in (2) is a Beta-normal (B-N) mixture.

The densities corresponding to A and G are illustrated in Figure 1.

Insert Figure 1 here

Pseudorandom numbers were generated using IMSL (1980) routines GGNPM for G and GGBTR for A. Each sample contained $n = 500$ observations with a predetermined number of observations, n_1, from component A, and the remaining $n - n_1$ from component G. Approximately ten samples were generated for each value of n_1 in the range 9 to 133, for a total of 1,275 samples in all.

This method of generating the mixed samples is different from that used in many previous studies where, in each sample, n_1 was generated as a pseudorandom outcome of n Bernoulli trials with success probability p. The component of variability due to differences in the generated distribution of n_1 is avoided in our study. Also, our method allows a more complete investigation of the relationship between estimates $\hat{p}$ and the sample fraction n_1/n.

Sample values were rounded to one decimal place to correspond with the slightly discrete nature of the Hb data. For each sample estimates were obtained using four different IPM rules, and NNML. For the IPM, sets $\{x_i\}$ were chosen with increments of $\cdot 1 = \sigma/10$ (I1) and $\cdot 2 = \sigma/5$ (I2). Each increment was used with a first reversal rule (RR1) and a second reversal rule (RR2) giving four procedures called RR1I1, RR1I2, RR2I1, and RR2I2. Convergence criteria used were $K_1 = K_2 = K_3 = 0$ except in the following situation. Due to the discreteness of the sample values, in some samples the convergence criteria were not met because oscillation between two sets of estimates was occurring, i.e., for some η, $(\hat{\mu}^{\nu}, \hat{\sigma}^{\nu}, \hat{p}^{\nu}) = (\hat{\mu}^{\nu+2}, \hat{\sigma}^{\nu+2}, \hat{p}^{\nu+2})$ but either (a) $\hat{\mu}^{\nu} \neq \hat{\mu}^{\nu+1}$ or (b) $\hat{\sigma}^{\nu} \neq \hat{\sigma}^{\nu+1}$, for $\nu > \eta$. The procedure adopted was to set $\hat{\mu}^{\nu*} = \frac{1}{2}(\hat{\mu}^{\eta+1} + \hat{\mu}^{\eta+2})$ if (a) held, or $\hat{\sigma}^{\nu*} = \frac{1}{2}(\hat{\sigma}^{\eta+1} + \hat{\sigma}^{\eta+2})$ if (b) held, and to redefine $K_1 = K_2 = \frac{1}{2}(x_i - x_{i-1})$, K_3 arbitrarily large. Convergence usually occurred at the next iteration. If not the procedure was halted (nonconvergence).

NNML was implemented using a combination of the gradient method (see, e.g., Hosmer, 1973b) and the Newton method (see, e.g., Dick and Bowden, 1973). It was found that 20 or 30 iterations using the gradient method followed by < 20 Newton iterations gave convergence in most samples. Nonconvergence of the NNML was declared if the estimate of p was not in (0,1) or if the estimate of one of the variances became negative. The convergence criterion used was that the length of the vector of first partial derivatives of the likelihood function be less than .01. Initial estimates were the ML estimates from the completely classified samples (Hosmer, 1973b), i.e., estimates of μ_1, σ_1, μ_2, σ_2, p, obtained knowing the component of origin of each observation in the sample. Thus the performance of NNML is examined under conditions which are favorable with respect to starting values, and the problem of different initial values leading to different estimates is avoided.

For a given method, sample means and variances of estimates (e.g., $\hat{p}$) for a given value of p were estimated using

$$E_p(\hat{p}) = E_{n_1|p}\{E(\hat{p}|n_1)\} = \sum_{n_1} E(\hat{p}|n_1)B(n,p,n_1)$$

and

$$Var_p(\hat{p}) = E_{n_1|p}\{Var(\hat{p}|n_1)\} + Var_{n_1|p}\{E(\hat{p}|n_1)\} \quad ,$$

where $E_{n_1|p}$, $Var_{n_1|p}$ represent moments with respect to the distribution of n_1 given p, with binomial probability mass function denoted $B(n,p,n_1)$. $E(\hat{p}|n_1)$, $Var(\hat{p}|n_1)$ were estimated by the sample mean and variance of the approximately ten estimates $\hat{p}$ for a given value of n_1. Samples where convergence did not occur were ignored in these computations.

Performance of each method was evaluated in terms of the sample means and variances of estimates for fixed p (see Tables 1, 2) and the sampling distribution of estimates over the range of values of n_1 (see Figures 2, 3, 4).

Insert Tables 1, 2 about here
Insert Figures 2, 3, 4 about here

Estimation of p is considered first. The relationship between the distribution of $\hat{p}$ and the sample fraction n_1/n is depicted in Figures 2, 3 and 4 for three of the methods. In these figures, the estimated 20th, 50th, and 80th percentiles (denoted A, B, and C, respectively) of the distribution of $\hat{p}$ for fixed n_1/n are plotted against the value of n_1/n. For most methods the distribution of $\hat{p}$ for fixed n_1 is skewed, so that bias is not a very meaningful criterion by which to evaluate performance. Nevertheless, sample means and standard deviations, estimated as described above, are presented in Table 1 for the different methods. As expected, $\hat{p}$ is positively biased for NNML, and increasingly negatively biased for RR1I1, the most conservative of the IPM rules. The other IPM rules have positive bias smaller than that of NNML. Sample variances are smallest for the conservative RR1I1 and generally largest for NNML and RR2I1.

Estimation of μ and σ is summarized in Table 2. Distribution of these estimates for fixed n_1 is again asymmetric. Inefficiency of the percentile (IPM) estimates of μ and σ is apparent. Bias is not consistently greater for NNML than for the percentile rules. This suggests that some method of ML estimation of μ and σ, in combination with (3.5) and (3.6) for estimation of p, e.g., as suggested in Section 4.2, may have better overall properties than either the IPM or NNML.

NNML estimates of the mean μ_A and standard deviation σ_A of the component A were extremely unreliable. For example, for $p = .20$ mean estimates of $-2.52 \pm .4293$ for $\mu_A = -2.7$, and $.92 \pm .1778$ for $\sigma_A = .80$, were obtained.

Bias in the results in Tables 1 and 2, due to ignoring samples where convergence did not occur, should be small, except for RR2I2, as convergence ratios were high (1271/1275 or 99.7% for RR1I1, 96.1% for RR1I2, 98.6% for RR2I1, 84.5% for RR2I2 and 93.7% for NNML). The low convergence ratio for RR2I2 is partly because overshoot (see Section 4.4) is more likely to result in nonconvergence with RR2I2 than with the other IPM rules. As a consequence, when convergence does occur, there is a tendency for the RR2I2 estimates to be more reliable. For $p > .05$, premature convergence of RR1I1 results in large negative biases for $\hat{p}$, $\hat{\mu}$ and positive bias for $\hat{\sigma}$.

6. COMPARISON OF THE IPM AND NNML FOR A NORMAL-NORMAL MIXTURE

In this section we change notation so as to distinguish between normal-normal (N-N) mixtures discussed here, and mixtures of the form (2) discussed earlier. For $i = 1, 2$ we denote by G_i the $N(\mu_i, \sigma_i^2)$ component cdf of a normal-normal mixture

$$F(x) = pG_1(x) + (1-p)G_2(x) \qquad 0<p<1, \quad -\infty \le x \le \infty \ . \tag{6}$$

In keeping with the situation described in Section 2, we assume p is small and $\mu_1 < \mu_2$. Hence G_1 and G_2 correspond to A and G, respectively, in (2). If $\mu_2 - \mu_1 > 3\sigma_1$, the correspondence is made closer as $G_1(x) \doteq 1$ for $x > \mu_2$, and (iv) of (2) holds approximately with $x_0 = \mu_2$.

For mixtures (6), it is known that the NNML estimates of p, μ_1 and σ_1 are unreliable if p is small (see, e.g., Hasselblad, 1966). There in such situations where estimation of p is important, the question arises as to whether a method such as the IPM, which concentrates on estimation of μ_2, σ_2 and p, may be better than NNML.

Although (b) is not a special case of (2), we have seen that (iv) of (2) holds approximately with $x_0 = \mu_2$, provided there is reasonable separation be-

tween the components, i.e., if $\mu_2 - \mu_1 \geq 3\sigma_1$. Defining $\lambda(x) = [F(x) - G_2(x)] / [1 - G_2(x)]$, then (3.4), on which the IPM is based, is also approximately satisfied.

The situation $|\mu_2 - \mu_1| = 3\min(\sigma_1, \sigma_2)$ has been referred to as "somewhat easy" for NNML (Fowlkes, 1979, see also Hosmer, 1973b), but previous simulation studies have not considered the effect of small p in this situation. We therefore carried out a Monte Carlo study to compare the IPM and NNML when the true model is (6) with $|\mu_2 - \mu_1| = 3\sigma_1$ and with emphasis on small p.

The parameter values used were $\mu_1 = -2.7$, $\sigma_1 = .9$, $\mu_2 = 0.0$, and $\sigma_2 = 1.0$. Thus G_2 is the same as G in the first or B-N study, also $\mu_1 = \mu_A$.

Samples of 500 observations were generated as in the B-N case, except that the n_1 observations from G_1 were obtained by transforming N(0,1) pseudorandom numbers, and a wider range of n_1 values (2 to 161) were included. An additional ten samples were generated for $n_1 = 10, 20, \cdots, 160$, giving a total of 1784 samples in all. Samples corresponding to n_1 values used in the B-N study were generated so that the $500 - n_1$ observations from G_2 were the same as the values from G in the B-N study. Thus differences in results for these samples could be directly attributed to the difference between A and G_1.

The four IPM procedures of Section 5.2 and NNML with the completely classified estimates (cce's) as initial estimates were applied to each sample. In addition for $n_1 = 2$ to 134, the RRIII estimates $\hat{p}$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and the first two sample moments were used to obtain estimates of μ_1 and σ_1 (see Dick and Bowden, 1973), and these five estimates were passed as initial estimates for NNML, the resulting procedure being denoted NNMLI.

The effect of the different initial estimates on NNML is considered first. Samples where RRIII did not converge and NNMLI was not applied are excluded leaving 1337 samples for $n_1 = 2$ to 134. The convergence ratio was lower for

NNML1, 1114/1337 or 83.3%, as compared to 94.6% for NNML. Both procedures converged in 1088 samples, and produced the same final estimates in all but 16 or 1.5% of these 1088 cases. As the cce's are not available as initial estimates in practice, obtaining initial estimates using RR1I1 as in NNML1 seems a reasonable approach when p is small. Results for NNML1 are not included in the following discussion and tables as they are so similar to those for NNML.

NNML estimates of μ_2, σ_2 are more efficient, and except for $p \leq .05$, generally less biased than the IPM estimates. This is expected because when the components are well separated, performance of NNML with respect to estimation of μ_2, σ_2 should not be substantially affected by p small (e.g., see Hasselblad, 1966, p. 441). Our chief concern is the effect on estimation of p.

For fixed n_1 the distribution of $\hat{p}$ is again asymmetric, with a long upper tail for all five methods (see Figures 5, 6, 7). The median of this distribution agrees well with the sample fraction n_1/n for NNML, even for $n_1/n < .04$ (Figure 7). The larger spread of $\hat{p}$ for NNML when n_1 is small is also noticeable in Figure 7. The negative bias of the conservative RR1I1 is increasingly apparent as n_1/n increases beyond $25/500 = .05$ in Figure 5.

Insert Figures 5, 6, 7 about here

Estimates of the sample means and standard deviations of $\hat{p}$ for the five methods are presented in Table 3 for a range of values of p. NNML is positively biased for small p ($p = .025, .05, .10$) but is generally superior to all IPM rules in terms of bias and precision for $p \geq .15$. For small p ($p \leq .05$) the conservative RR1I1 is the best method. The negative bias of all IPM rules becomes increasingly pronounced as p increases beyond .15.

Insert Table 3 here

Convergence ratios for the 1384 samples were generally similar to those for the Beta-normal mixture (98.9% for RR1I1, 96.5% for RR1I2, 97.6% for RR2I1, 85.4% for RR2I2 and 95.0% for NNML).

7. DISCUSSION

Tables 1, 2 and 3 suggest that when estimation of p is important there are situations where the IPM may be more appropriate than NNML. These occur when A is markedly nonnormal for any p, or when A is normal but p is very small (say, $p \leq .05$ when $n = 500$, or $np < 25$-30 for larger n).

Differences between performance for each method for the two types of mixture are evident in Table 1. As stated earlier, this table is calculated from samples generated so that the values from G in the B-N mixture were the same as the values from G_2 in the N-N mixture. The effect of nonnormality of A on NNML is evident in the increased bias and variance of $\hat{p}$, especially for $p = .20$. Precision tends to be better for IPM rules in the N-N situation, probably because in some sense there is better separation between the two components in the N-N case (i.e., $G_1(x) < A(x)$ for $-2.7 < x < -1.5$).

Results for estimation of $\hat{\mu}$, $\hat{\sigma}$ in Table 2, and for $\hat{\mu}_2$, $\hat{\sigma}_2$ not shown, suggest that a combination of ML estimation of $\hat{\mu}$, $\hat{\sigma}$ (or $\hat{\mu}_2$, $\hat{\sigma}_2$) and a reversal rule based on $\hat{\lambda}$ for estimation of p may be the best approach when p is small. One such procedure was outlined in Section 4.2. The use of an IPM rule to obtain initial estimates for NNML when p is small seems worth pursuing.

The discreteness of the simulated values (due to rounding to one decimal place) is typical of the accuracy with which Hb values, and other measures of blood parameters, are reported. For data of this type there are natural choices for the increments $(x_i - x_{i-1})$ and the convergence criteria K_i, $i = 1, 2, 3$. This reduces some of the subjectivity involved in applying the IPM, but the problem of optimal rules still remains. Empirical evidence suggests that it may be

worthwhile investigating further the use of iterative percentile methods such as those suggested here for mixture problems when p is small.

8. EXAMPLE

The IPM RR1 rule and the NNML method were applied to Hb values for women between 18 and 44 years of age from the First Health and Nutrition Examination Survey (1971-1974). A detailed discussion of this data set is contained in Meyers (1978). Values for 3,074 white and 742 black women were analyzed separately. Initial estimates for NNML were obtained using Hald's truncation procedure as described in Hasselblad (1966). Values were grouped forming classes of length 0.3 and iteration was by the gradient method as described by Hasselblad (1966). The conservative RRlIl rule with $K_1 = K_2 = K_3 = 0$ and $(x_i - x_{i-1}) = \cdot 1$ was used because a plot of the sample cdf on normal probability paper suggested that the non-anemic component was highly predominant (see Meyers, 1978).

Results for the two methods (RRlIl and NNML) are presented in Table 4. Agreement between the two methods is remarkably close. Standard errors for the RRlIl estimates were verified by a small simulation of an N-N mixture with parameter values equal to the NNML estimates, for both samples of women. In spite of the fact that Appendix formulae (A.1) and (A.2) are based on the assumption that $\hat{\mu}$, $\hat{\sigma}$ are continuous, whereas for Hb data they are discrete, good agreement was obtained. There is as yet no satisfactory method for obtaining a standard error for $\hat{p}_{RRlIl}$. Simulation results suggested a standard error of the order of $\cdot 3\%$ for white women, and $1 \cdot 7\%$ for black women. Note that there appears to be reasonable separation between the anemic and non-anemic components, i.e., $\hat{\mu}_2 - \hat{\mu}_1 > 3\sigma_1$ for white and black women.

Insert Table 4 here

Finally, it is of interest to compare the prevalence estimates in Table 4 with estimates obtained using the WHO criterion which classifies all women with Hb values less than 12g/dl as anemic. The latter estimates are $4.5 \pm .4\%$ for white women and $20.4 \pm 1.5\%$ for black women (see Meyers, 1978).

With the model (2) and $x_0 = 12$g/dl by the WHO definition, and using the RRlIl estimates of $\hat{\mu}$, $\hat{\sigma}$, and Appendix formula (A.6), we have

$$\hat{\lambda}(12) = \begin{cases} 0.9 \pm .6\% & \text{for white women} \\ 2.9 \pm 2.4\% & \text{for black women} \end{cases} .$$

Allowing for the fact that the distribution of $\hat{\lambda}$ is asymmetric with a long upper tail, these results suggest that the WHO criterion leads to overestimates of the prevalence of anemic women in these samples. For further discussion of this issue, see Meyers (1978).

APPENDIX

Asymptotic variances and covariances for IPM estimates are presented below. These will be appropriate in practice for small p (i.e., $p \leq 5\%$ or so).

Let $\varphi(\cdot)$ represent the standard normal density function.

$$\mathrm{Var}(\hat{\mu}) \doteq \frac{F(\mu)[1-F(\mu)]}{N[(1-p)\varphi(0)/\sigma]^2} = \frac{\sigma^2}{N}\,\frac{F(\mu)[1-F(\mu)]}{(1-p)^2\varphi(0)^2} \quad , \tag{A.1}$$

$$\mathrm{Var}(\hat{\sigma}) \doteq \frac{\sigma^2}{N}\left\{\frac{F(\mu)[1-F(\mu)]}{(1-p)^2\varphi(0)^2} + \frac{F(\mu+\sigma)[1-F(\mu+\sigma)]}{(1-p)^2\varphi(1)^2} - \frac{2F(\mu)[1-F(\mu+\sigma)]}{(1-p)^2\varphi(0)\varphi(1)}\right\} , \tag{A.2}$$

$$\mathrm{Cov}(\hat{\mu},\hat{\sigma}) \doteq \frac{\sigma^2}{N}\,\frac{F(\mu)[1-F(\mu+\sigma)]}{(1-p)^2\varphi(0)\varphi(1)} \tag{A.3}$$

and

$$\mathrm{Var}\hat{G}(x) \doteq \frac{1}{\sigma^2}\left[\varphi\left(\frac{x-\mu}{\sigma}\right)\right]^2\left\{\mathrm{Var}(\hat{\mu}) + \left(\frac{x-\mu}{\sigma}\right)^2\mathrm{Var}(\hat{\sigma}) + 2\left(\frac{x-\mu}{\sigma}\right)\mathrm{Cov}(\hat{\mu},\hat{\sigma})\right\} \quad . \tag{A.4}$$

Estimates of (A.1), (A.2), (A.3) and (A.4) are obtained in terms of $\hat{p}$, $\hat{\mu}$, $\hat{\sigma}$ by noting that

$$F(\mu) = p + (1-p)(.5) \qquad \text{and} \qquad F(\mu+\sigma) = p + (1-p)(.8413) \quad .$$

$$\mathrm{Cov}[\hat{F}(x_1),\hat{G}(x_2)] \doteq -\frac{1}{\sigma}\varphi\left(\frac{x_2-\mu}{\sigma}\right)\left\{\mathrm{Cov}[\hat{F}(x_1),\hat{\mu}] + \left(\frac{x_2-\mu}{\sigma}\right)\mathrm{Cov}[\hat{F}(x_1)\hat{\sigma}]\right\} \quad . \tag{A.5}$$

This involves covariances of the form

$$\mathrm{Cov}[\hat{F}(x),\hat{\mu}] \doteq \begin{cases} \dfrac{-F(x)[1-F(\hat{\mu})]}{N(1-p)\varphi(0)/\sigma} & \text{if } x < \mu \\[2ex] \dfrac{-F(\hat{\mu})[1-F(x)]}{N(1-p)\varphi(0)/\sigma} & \text{if } x > \mu \end{cases} .$$

$$\mathrm{Var}\hat{\lambda}(x) \doteq \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{-2}\{\mathrm{Var}\hat{F}(x) + [1 - \lambda(x)]^2\mathrm{Var}\hat{G}(x) - 2[1 - \lambda(x)]\mathrm{Cov}[\hat{F}(x), \hat{G}(x)]\} \qquad (A.6)$$

and

$$\mathrm{Cov}[\hat{\lambda}(x_1), \hat{\lambda}(x_2)] \doteq \left[1 - \Phi\left(\frac{x_1-\mu}{\sigma}\right)\right]^{-1}\left[1 - \Phi\left(\frac{x_2-\mu}{\sigma}\right)\right]^{-1}\{\mathrm{Cov}[\hat{F}(x_1), \hat{F}(x_2)] - [1 - \lambda(x_2)]\mathrm{Cov}[\hat{F}(x_1), \hat{G}(x_2)] - [1 - \lambda(x_1)]\mathrm{Cov}[\hat{F}(x_2), \hat{G}(x_1)] + [1 - \lambda(x_1)][1 - \lambda(x_2)]\mathrm{Cov}[\hat{G}(x_1), \hat{G}(x_2)]\} \quad , \qquad (A.7)$$

where

$$\mathrm{Cov}[\hat{G}(x_1), \hat{G}(x_2)] \doteq \frac{1}{\sigma^2}\varphi\left(\frac{x_1-\mu}{\sigma}\right)\varphi\left(\frac{x_2-\mu}{\sigma}\right)\left\{\mathrm{Var}(\hat{\mu}) + \left(\frac{x_2-\mu}{\sigma} + \frac{x_1-\mu}{\sigma}\right)\mathrm{Cov}(\hat{\mu}, \hat{\sigma}) + \left(\frac{x_1-\mu}{\sigma}\right)\left(\frac{x_2-\mu}{\sigma}\right)\mathrm{Var}(\hat{\sigma})\right\} \quad .$$

Table 1. Comparison of IPM and NNML for Beta-Normal (B-N) and Normal-Normal (N-N) Mixtures, and Different Values of p

Estimated Sample Means and Standard Deviations of $\hat{p}$

	p = .05		p = .10		p = .15		p = .20	
Method	B-N	N-N	B-N	N-N	B-N	N-N	B-N	N-N
RR1I1	.054 (.0537)	.052 (.0495)	.093 (.0623)	.089 (.0524)	.130 (.0675)	.123 (.0494)	.177 (.0766)	.168 (.0607)
RR1I2	.066 (.070)	.065 (.0722)	.112 (.0823)	.109 (.0772)	.161 (.0845)	.154 (.0833)	.216 (.0805)	.205 (.0866)
RR2I1	.067 (.0647)	.061 (.0590)	.110 (.069)	.105 (.0685)	.156 (.0773)	.146 (.0677)	.218 (.0865	.196 (.0785)
RR2I2	.069 (.0575)	.062 (.0536)	.112 (.0720)	.113 (.0805)	.157 (.0756)	.148 (.0675)	.208 (.0723)	.198 (.0746)
NNML	.078 (.1009)	.089 (.1071)	.1274 (.1043)	.126 (.0829)	.168 (.0744)	.162 (.0639)	.235 (.0877)	.208 (.0576)
Values of n_1	9 – 45		27 – 76		47 – 105		69 – 133	
No. of samples	385		513		600		656	

Table 2. Comparison of IPM and NNML Estimates of $\mu = 0$ and $\sigma = 1.0$ for the Beta-Normal Mixture and Different Values of p

Estimated Sample Means and Standard Deviations of $\hat{\mu}$ and $\hat{\sigma}$

Method	p = .05		p = .10		p = .15		p = .20	
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
RR1I1	-.01 (.1012)	1.00 (.0970)	-.02 (.1242)	1.01 (.1060)	-.05 (.1418)	1.02 (.1163)	-.05 (.1565)	1.03 (.1238)
RR1I2	.00 (.1188)	1.00 (.0980)	-.00 (.1464)	1.01 (.1120)	-.01 (.1610)	1.01 (.1200)	.01 (.1552)	1.00 (.1208)
RR2I1	.00 (.1155)	1.00 (.0984)	-.00 (.1335)	1.00 (.1059)	-.02 (.1541)	1.01 (.1180)	.01 (.1707)	1.00 (.1278)
RR2I2	.00 (.0992)	1.00 (.0975)	-.01 (.1271)	1.02 (.0954)	-.03 (.1486)	1.02 (.1163)	-.02 (.1445)	1.02 (.1134)
NNML	.02 (.1169)	.99 (.0706)	.02 (.1278)	1.00 (.0873)	.01 (.1277)	1.00 (.0810)	.04 (.1429)	.98 (.0946)
Values of n_1	9 – 45		27 – 76		47 – 105		69 – 133	
No. of samples	385		513		600		656	

Table 3. Comparison of IPM and NNML for the Normal-Normal Mixture for Different Values of p

Estimated Sample Means and Standard Deviations of $\hat{p}$

Method	p = .025	p = .05	p = .10	p = .15	p = .20	p = .25
RR1I1	.030 (.0363)	.051 (.0489)	.088 (.0515)	.122 (.0486)	.168 (.0599)	.207 (.0663)
RR1I2	.039 (.0505)	.064 (.0721)	.109 (.0775)	.153 (.0830)	.205 (.0851)	.242 (.0748)
RR2I1	.044 (.0608)	.060 (.0587)	.104 (.0685)	.145 (.0669)	.196 (.0773)	.240 (.0716)
RR2I2	.045 (.0594)	.062 (.0537)	.114 (.0809)	.147 (.0669)	.197 (.0725)	.241 (.0713)
NNML	.046 (.0685)	.088 (.1059)	.126 (.0824)	.162 (.0643)	.208 (.0574)	.259 (.0642)
Values of n_1	2 – 28	9 – 45	27 – 76	47 – 105	69 – 133	91 – 161
No. of samples	297	425	563	660	726	781

Table 4. Results for Analysis of Hb Values (g/dl) for Mixed Sample of Anemic and Non-Anemic Women:

Estimated Prevalence of Anemia ($\hat{p}$) and Hb means ($\hat{\mu}$) and Standard Deviations ($\hat{\sigma}$)

Estimation Method	Sample of 3074 white women: Anemics		Non-anemics		Prevalence of anemia (%)
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{p}$(%)
NNML	9.70 ± .420[a]	1.19 ± .270	13.80 ± .019	1.03 ± .014	1.00 ± .271
RR1I1			13.8 ± .02[b]	1.0 ± .03	1.1
	Sample of 742 black women: Anemics		Non-anemics		Prevalence of anemia (%)
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{p}$(%)
NNML	9.47 ± .466	0.88 ± .273	13.12 ± .050	1.16 ± .039	2.73 ± 1.156
RR1I1			13.1 ± .06	1.2 ± .07	2.5

a Estimated asymptotic standard error of NNML estimate.

b Standard error for RR1I1 estimate, obtained as in Appendix.

Figure 1. Component Densities a(x), g(x) of the Beta-Normal Mixture.

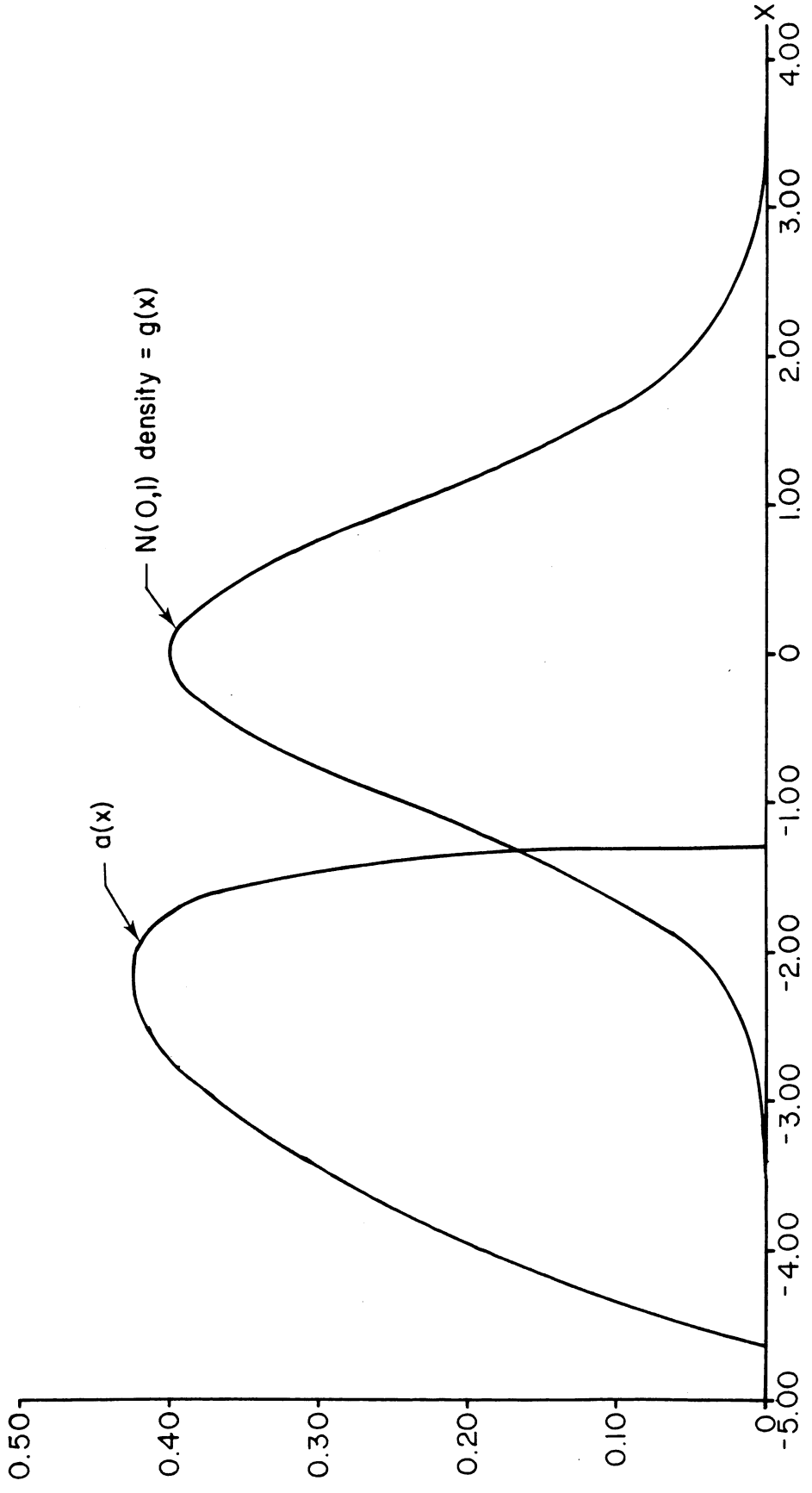

N(0,1) density = g(x)
a(x)
0.50
0.40
0.30
0.20
0.10
0
-5.00
-4.00
-3.00
-2.00
-1.00
0
1.00
2.00
3.00
4.00
X

Figure 2. Sampling Distribution of $\hat{p}_{RR1I1}$ for Varying n_1/n, B-N Case.

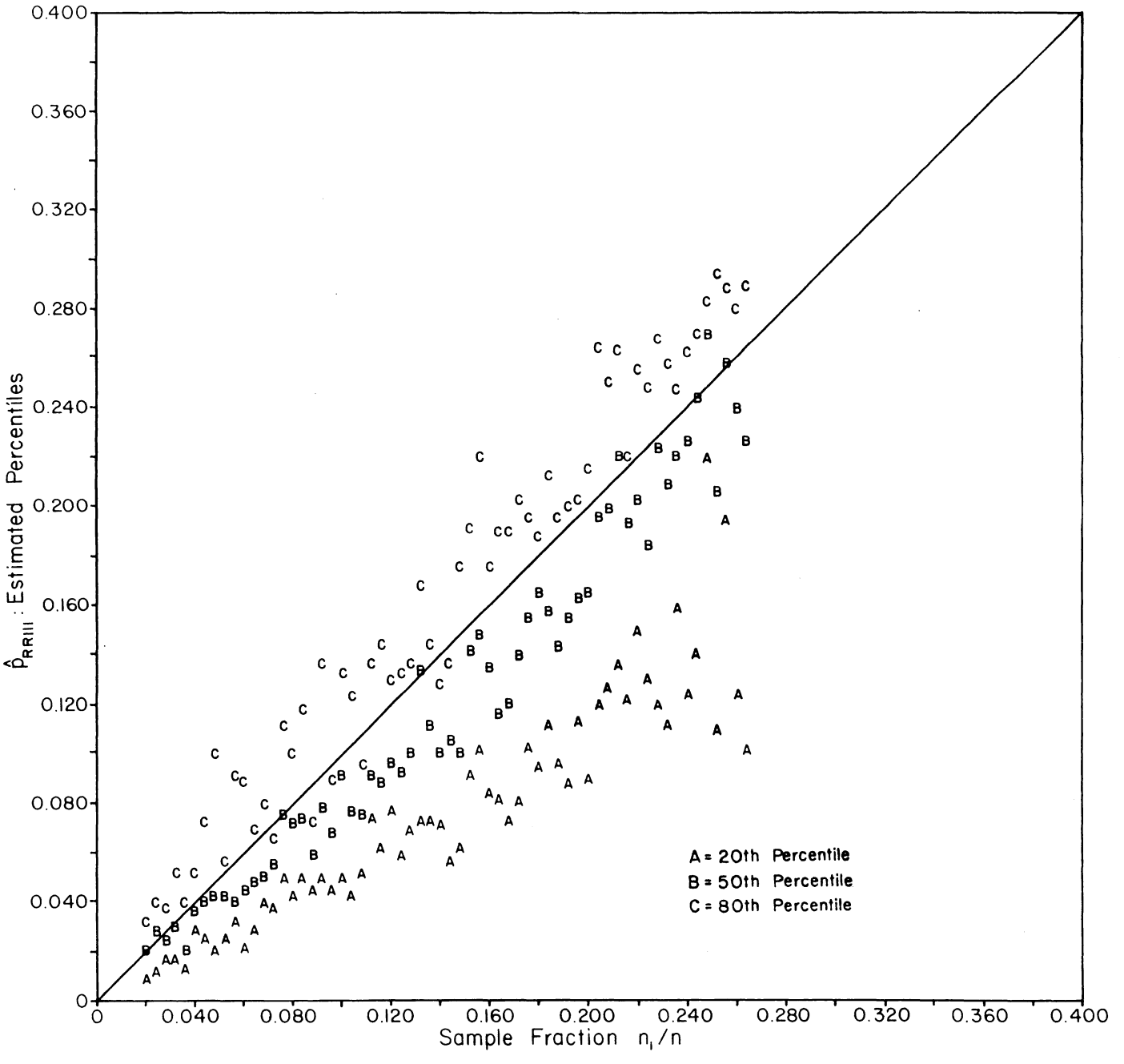

0.400
0.360
0.320
0.280
0.240
0.200
0.160
0.120
0.080
0.040
0
$\hat{p}_{RRIII}$: Estimated Percentiles
0
0.040
0.080
0.120
0.160
0.200
0.240
0.280
0.320
0.360
0.400
Sample Fraction n_1/n
A = 20th Percentile
B = 50th Percentile
C = 80th Percentile

Figure 3. Sampling Distribution of $\hat{p}_{RR2I2}$ for Varying n_1/n, B-N Case.

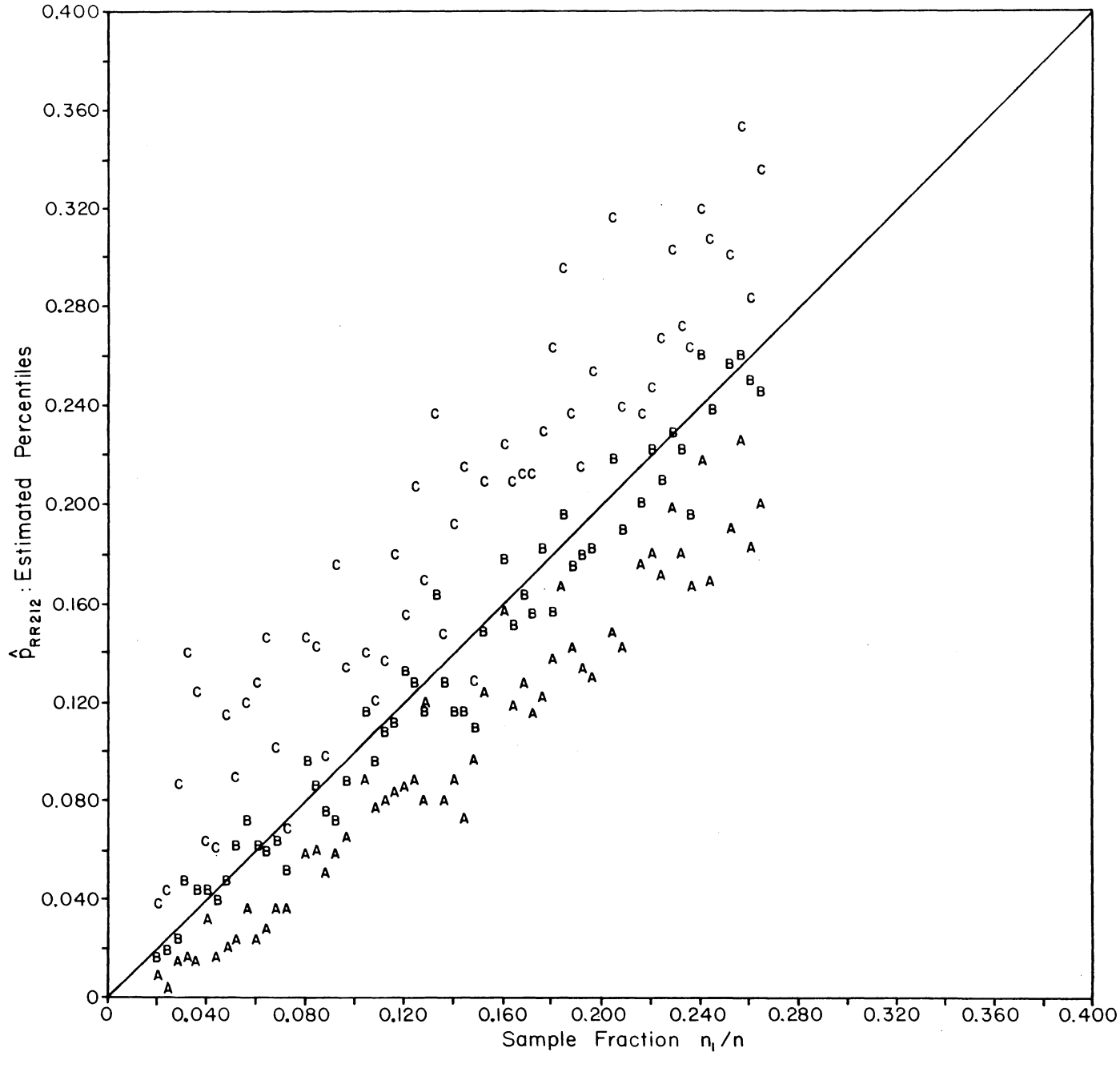

0.400
0.360
0.320
0.280
0.240
0.200
0.160
0.120
0.080
0.040
0
$\hat{p}_{RR212}$: Estimated Percentiles
0
0.040
0.080
0.120
0.160
0.200
0.240
0.280
0.320
0.360
0.400
Sample Fraction n_i/n

Figure 4. Sampling Distribution of $\hat{p}_{NNML}$ for Varying n_1/n, B-N Case.

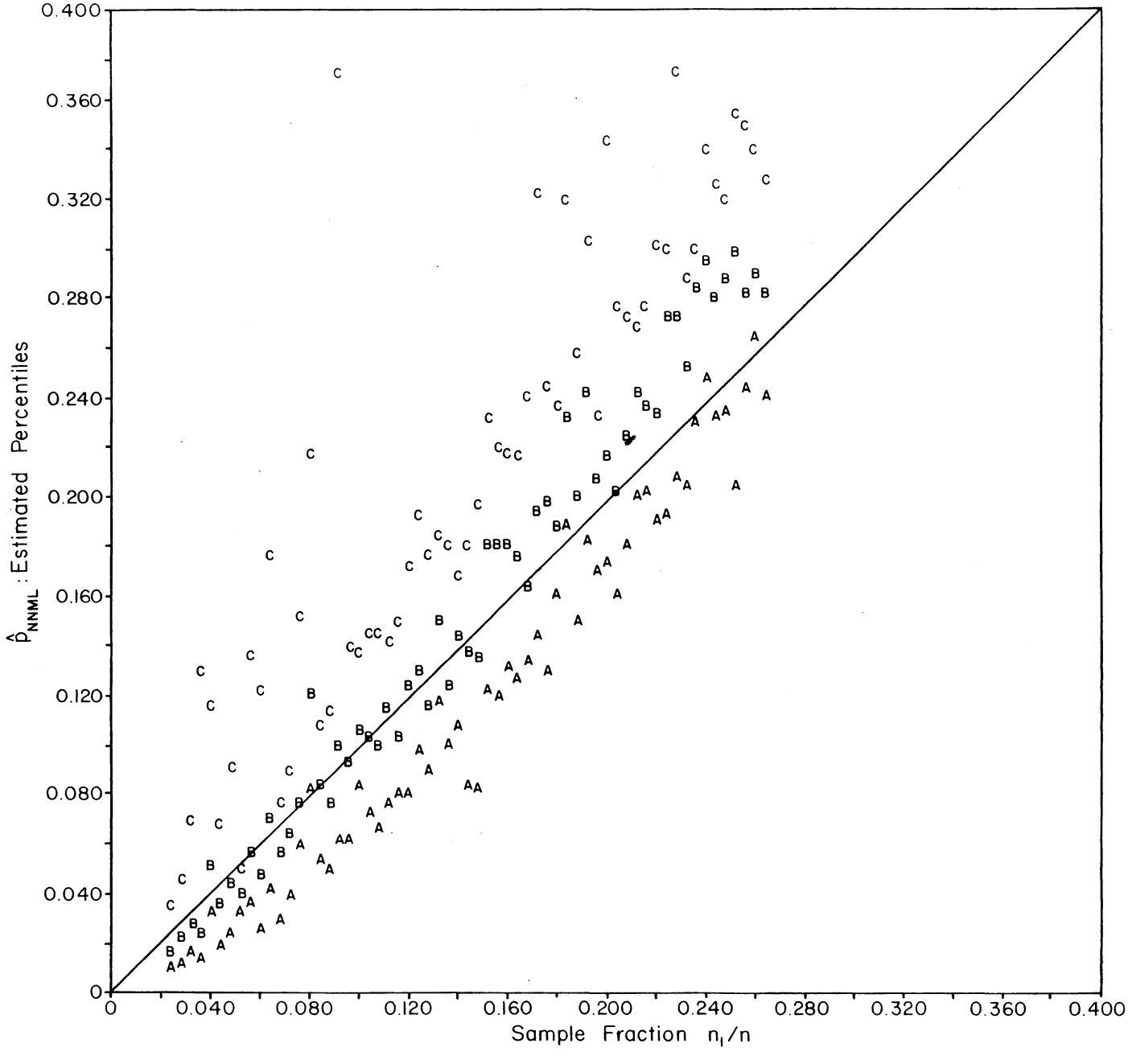

0.400
0.360
0.320
0.280
0.240
0.200
0.160
0.120
0.080
0.040
0
$\hat{p}_{NNML}$: Estimated Percentiles
0
0.040
0.080
0.120
0.160
0.200
0.240
0.280
0.320
0.360
0.400
Sample Fraction n_1/n

Figure 5. Sampling Distribution of $\hat{p}_{RRlll}$ for Varying n_1/n, N-N Case.

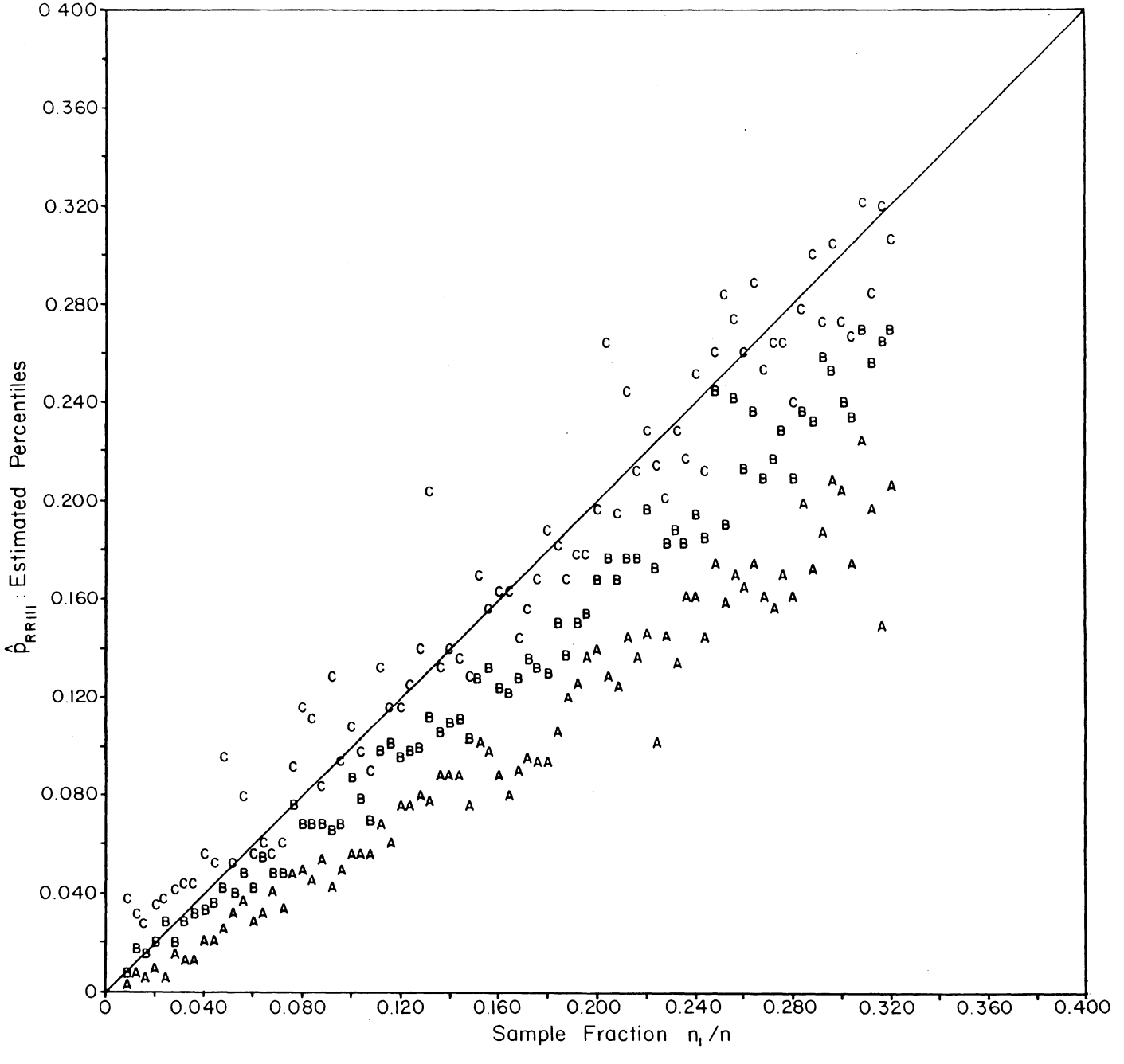

0 400
0.360
0.320
0.280
0.240
0.200
0.160
0.120
0.080
0.040
0
$\hat{p}_{RRIII}$: Estimated Percentiles
0
0.040
0.080
0.120
0.160
0.200
0.240
0.280
0.320
0.360
0.400
Sample Fraction n_1/n

Figure 6. Sampling Distribution of $\hat{p}_{RR2I2}$ for Varying n_1/n, N-N Case.

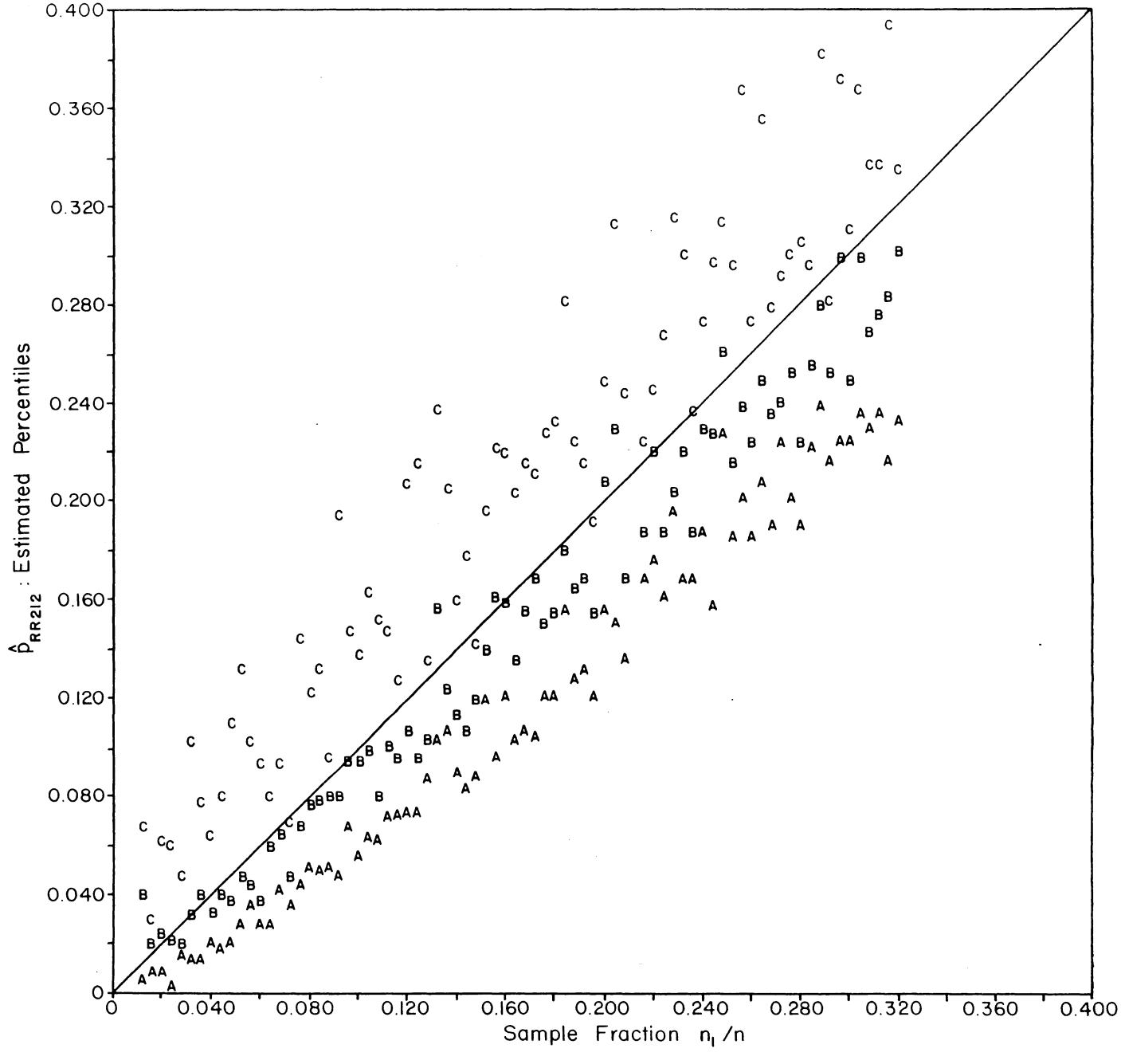
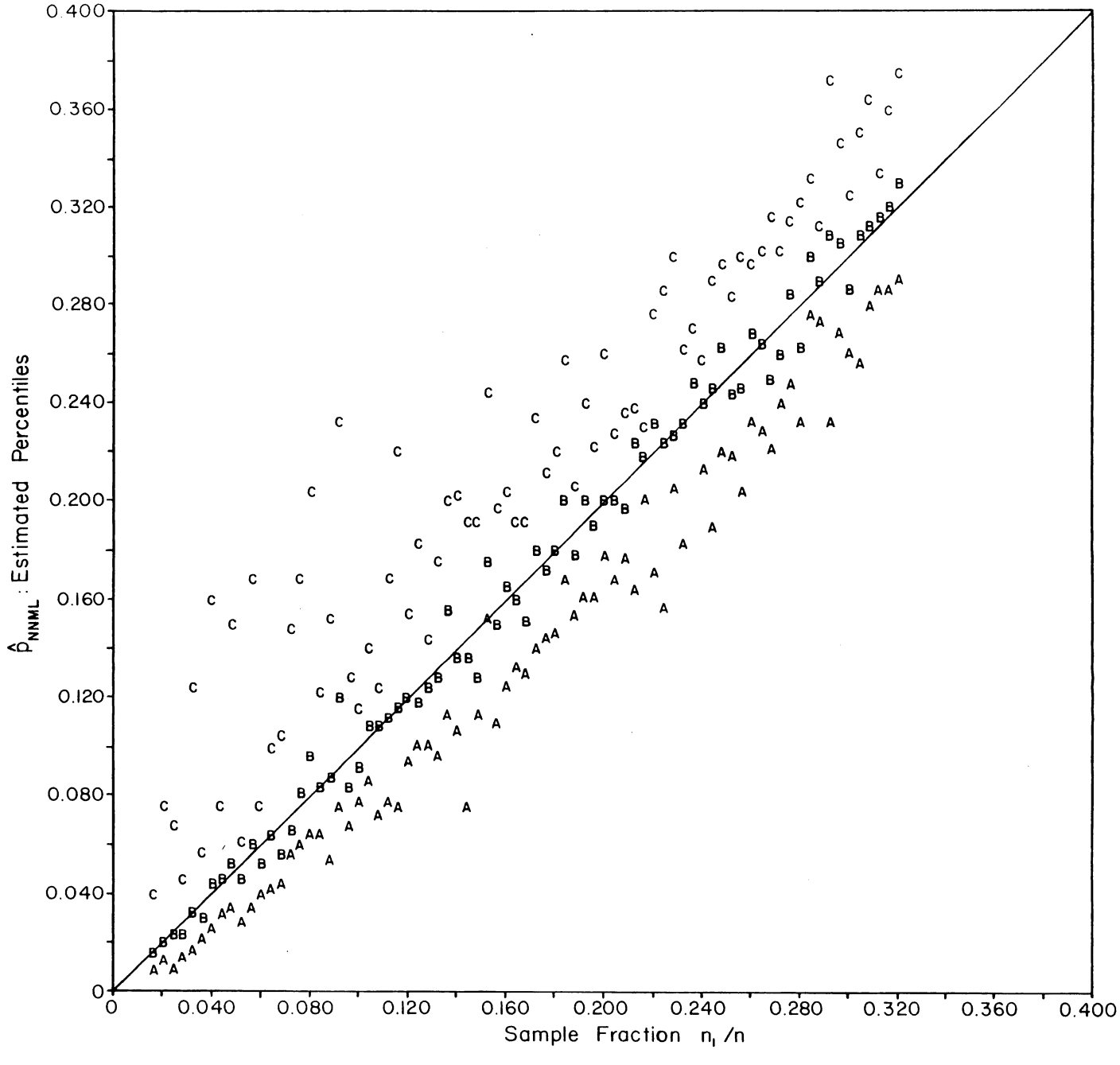

0.400
0.360
0.320
0.280
0.240
0.200
0.160
0.120
0.080
0.040
0
$\hat{P}_{RR212}$: Estimated Percentiles
0
0.040
0.080
0.120
0.160
0.200
0.240
0.280
0.320
0.360
0.400
Sample Fraction n_1/n

Figure 7. Sampling Distribution of $\hat{p}_{NNML}$ for Varying n_1/n, N-N Case.


0.400
0.360
0.320
0.280
0.240
0.200
0.160
0.120
0.080
0.040
0
$\hat{p}_{NNML}$: Estimated Percentiles
0
0.040
0.080
0.120
0.160
0.200
0.240
0.280
0.320
0.360
0.400
Sample Fraction n_1/n

REFERENCES

BECKTEL, J. M. (1970), "Simplified Estimation of Normal Ranges from Routine Laboratory Data," Clinica Chimica Acta, 28, 119-125.

BINDER, DAVID A. (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 746-747.

BRYANT, PETER (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regression' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 748-749.

CLARK, VIRGINIA A., CHAPMAN, J., COULSON, A. and HASSELBLAD, V. (1968), "Dividing Blood Pressures from the Los Angeles Heart Study into Two Normal Distributions," The Johns Hopkins Medical Journal, 122, 77-83.

CLARKE, B. R. and HEATHCOTE, C. B. (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey." Journal of the American Statistical Association, 73, 749-750.

COHEN, A. CLIFFORD (1950), "Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples," Annals of Mathematical Statistics, 21, 557-569.

COOK, JAMES D., ALVARADO, J., GUTNISKY, A., JAMRA, M., LABARDINI, J., LAYRISSE, M., LINARES, J., LORÍA, A., MASPES, V., RESTREPO, A., REYNAFARJE, C., SÁNCHEZ-MEDAL, L., VÉLEZ, H. and VITERI, F. (1971), "Nutritional Deficiency and Anemia in Latin America: A Collaborative Study," Blood, 38, 591-603.

DAY, N. E. (1969), "Estimating the Components of a Mixture of Normal Distributions," Biometrika, 56, 463-474.

DICK, NATHAN P. and BOWDEN, D. C. (1973), "Maximum Likelihood Estimation for Mixtures of Two Normal Distributions," Biometrics, 29, 781-790.

ELSTON, R. C., NAMBOODIRI, K. K., NINO, H. V. and POLLITZER, W. S. (1974), "Studies on Blood and Urine Glucose in Seminole Indians: Indications for Segregation of a Major Gene," American Journal of Human Genetics, 26, 13-34.

FOWLKES, EDWARD B. (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 747-748.

_____ (1979), "Some Methods for Studying the Mixture of Two Normal (Lognormal) Distributions," Journal of the American Statistical Association, 74, 561-575.

FRYER, J. G. and ROBERTSON, C. A. (1972), "A Comparison of Some Methods for Estimating Mixed Normal Distributions," Biometrika, 59, 639-648.

GARBY, LARS, IRNELL, L. and WERNER, I. (1969), "Iron Deficiency in Women of Fertile Age in a Swedish Community. III. Estimation of Prevalence Based on Response to Iron Supplementation," Acta Medica Scandinavica, 185, 113-117.

GINDLER, E. MELVIN (1970), "Calculation of Normal Ranges by Methods Used for Resolution of Overlapping Gaussian Distributions," Clinical Chemistry, 16, 124-128.

HARTLEY, MICHAEL J. (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 738-741.

HASSELBLAD, VICTOR (1966), "Estimation of Parameters for a Mixture of Normal Distributions," Technometrics, 8, 431-444.

HOSMER, DAVID W., Jr. (1973a), "A Comparison of Iterative Maximum Likelihood Estimates of the Parameters of a Mixture of Two Normal Distributions Under Three Different Types of Sample," Biometrics, 29, 761-770.

_____ (1973b), "On MLE of the Parameters of a Mixture of Two Normal Distributions when the Sample Size is Small," Communications in Statistics, 1, 217-227.

_____ (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 741-744.

IMSL (1980), International Mathematical and Statistical Libraries, Inc. Library 1, Edition 7.1, Houston, Texas.

JOHNSON, NORMAN L. (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 741-744.

KENDALL, MAURICE G. and STUART, ALAN (1977), The Advanced Theory of Statistics, Vol. 1, New York: MacMillan.

KIEFER, JACK and WOLFOWITZ, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," Annals of Mathematical Statistics, 27, 887-906.

KIEFER, NICHOLAS M. (1978a), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions' by R. E. Quandt and J. B. Ramsey," Journal of the American Statistical Association, 73, 744-745.

_____ (1978b), "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model," Econometrica, 46, 427-434.

MACDONALD, P. D. M. and PITCHER, T. J. (1979), "Age Groups from Size-Frequency Data: A Versatile and Efficient Method of Analyzing Distribution Mixtures," Journal of the Fisheries Research Board of Canada, 36, 987-1001.

MARTIN, H. F., GUDZINOWICZ, B. J. and FANGER, H. (1975), Normal Values in Clinical Chemistry – A Guide to Statistical Analysis of Laboratory Data, New York: Marcel Dekker.

MEYERS, LINDA D. (1978), "Definition, Prevalence and Correlates of Iron Deficiency Anemia in Black and White American Women – An Epidemiologic Analysis," Ph.D. thesis, Cornell University, Ithaca, New York.

NEUMANN, GEORGE J. (1968), "The Determination of Normal Ranges from Routine Laboratory Data," Clinical Chemistry, 14, 979-988.

PEARSON, KARL (1894), "Contributions to the Mathematical Theory of Evolution," Philosophical Transactions of the Royal Society, Ser. A, 185, 71-110.

QUANDT, RICHARD E. (1972), "A New Approach to Estimating Switching Regressions," Journal of the American Statistical Association, 67, 306-310.

_____ and RAMSEY, JAMES B. (1978), "Estimating Mixtures of Normal Distributions and Switching Regressions," Journal of the American Statistical Association, 73, 730-738.

TAN, W. Y. and CHANG, W. C. (1972), "Some Comparison of the Method of Moments and the Method of Maximum Likelihood in Estimating Parameters of a Mixture of Two Normal Densities," Journal of the American Statistical Association, 67, 702-708.

WHO (1968), Nutritional Anemias, Report of a World Health Organization Scientific Group, WHO Technical Report Series No. 405.